

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 11-272666

(43)Date of publication of application : 08.10.1999

(51)Int.Cl.

G06F 17/27

G06F 3/00

(21)Application number : 10-072727

(71)Applicant : FUJITSU LTD

(22)Date of filing : 20.03.1998

(72)Inventor : FUJIKAWA YASUYUKI

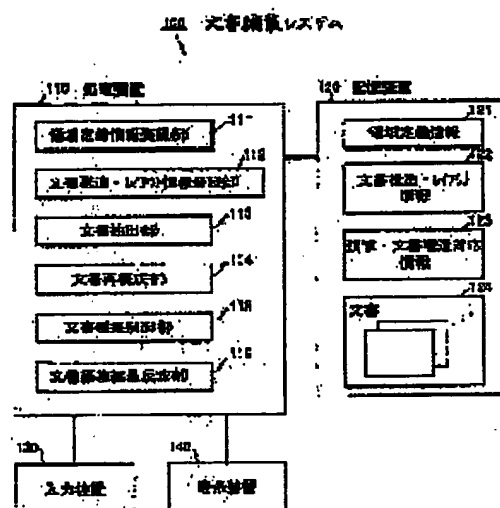
(54) SYSTEM AND METHOD FOR EDITING DOCUMENT AND RECORD MEDIUM

(57)Abstract:

PROBLEM TO BE SOLVED: To obtain the document editing system which reconstitute areas extracted from plural documents with specific pattern and displays in the state that a user can edit them.

SOLUTION: A set of generally usable start and end patterns and document structure information are used to extract areas between places corresponding to the set of the patterns from plural documents, and the results are put together and reconstituted according to the document structure information and layout information 122 and displayed in the state that the user can edit them. The areas edited by the user are written back to the documents of the extraction sources.

Further, the document structure information and layout information 122 are registered and held together; and the document structure information includes the information on the hierarchical structure (inclusion relation) of areas in the documents, appearance order, the presence of the repetition part, etc., and the layout information includes the information on the display positions, fonts styles, and font sizes of the respective areas, the arrangement and character color of character strings, etc.



LEGAL STATUS

[Date of request for examination] 20.11.2000

[Date of sending the examiner's decision of rejection] 16.08.2005

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number] 3737629

[Date of registration] 04.11.2005

[Number of appeal against examiner's decision of rejection] 2005-17682

[Date of requesting appeal against examiner's decision of rejection] 15.09.2005

[Date of extinction of right]

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平11-272666

(43) 公開日 平成11年(1999)10月8日

(51) Int.Cl.⁶

G 0 6 F 17/27
3/00

識別記号

6 5 5

F I

G 0 6 F 15/20
3/00

6 5 0 E
6 5 5 A

審査請求 未請求 請求項の数12 O L (全 22 頁)

(21) 出願番号 特願平10-72727

(22) 出願日 平成10年(1998) 3月20日

(71) 出願人 000005223

富士通株式会社

神奈川県川崎市中原区上小田中4丁目1番
1号

(72) 発明者 藤川 泰之

神奈川県川崎市中原区上小田中4丁目1番
1号 富士通株式会社内

(74) 代理人 弁理士 大曾 義之 (外1名)

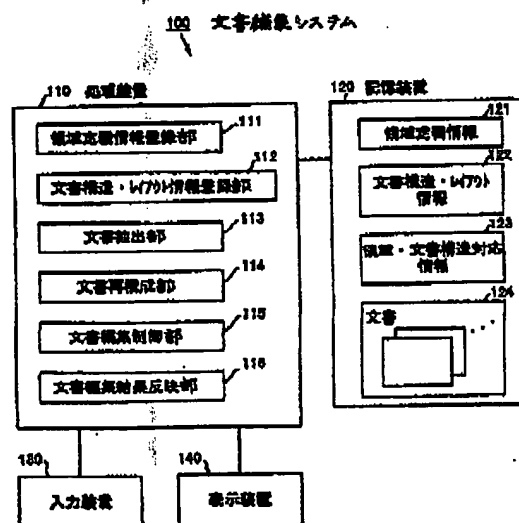
(54) 【発明の名称】 文書編集システム、方法、及び記録媒体

(57) 【要約】

【課題】 複数の文書から所定のパターンによって抽出された領域を再構成して、ユーザが編集可能な状態で表示させる文書編集システムを提供する。

【解決手段】 汎用的な開始パターン及び終了パターンからなるパターンの組と文書構造情報を用いて、複数の文書から前記パターンの組に該当する個所の間の領域をそれぞれ抽出し、結果を前記文書構造情報とレイアウト情報に従って集約して再構成し、ユーザが編集可能な状態で表示する。ユーザにより編集された領域は抽出元の文書に書き戻される。更に、前記文書構造情報と前記レイアウト情報は一体的に登録、保持され、前記文書構造情報は前記文書内の領域の階層構造(包含関係)、出現順序、及び繰り返しの有無等の文書構造に関する情報を含み、前記レイアウト情報は、前記各領域の表示位置、フォント・スタイル、フォント・サイズ、文字列の配置、及び文字色等のレイアウト情報を含む。

本発明の文書編集システムの
一実施の形態のシステム構成を示すブロック図



【特許請求の範囲】

【請求項1】 少なくとも1つの文書から任意の領域を抽出して、該抽出された領域を編集可能な状態で表示する文書編集システムであって、

各文書から所望の領域を抽出するために使用される領域定義情報をパターンとして登録する領域定義情報登録手段と、

各文書内の領域の構造に関する文書構造情報、及び該各領域を再構成するために使用されるレイアウト情報を登録する文書構造・レイアウト情報登録手段と、

該記領域定義登録手段によって登録された領域定義情報と、該文書構造・レイアウト情報登録手段によって登録された文書構造情報を使用して領域を抽出する文書抽出手段と、

該文書構造・レイアウト情報登録手段によって登録された該文書構造情報及び該レイアウト情報を使用して、該抽出された少なくとも1つの領域を再構成して表示する文書再構成手段と、

を有することを特徴とする文書編集システム。

【請求項2】 前記再構成して表示された少なくとも1つの領域を編集できるように制御する文書編集制御手段を有すること、

を特徴とする請求項1に記載の文書編集システム。

【請求項3】 前記文書編集制御手段が更に、前記再構成して表示された少なくとも1つの領域に関する編集機能を、前記文書から抽出された領域の内容のみに制限するように制御すること、

を特徴とする請求項2に記載の文書編集システム。

【請求項4】 前記抽出された少なくとも1つの領域が編集された場合、その編集結果を該領域の抽出元の文書に反映させる文書編集結果反映手段を有すること、

を特徴とする請求項2に記載の文書編集システム。

【請求項5】 前記領域定義情報登録手段によって登録される領域定義情報が、前記所望の領域の直前の箇所を特定する第1のパターンを含み、

前記文書構造・レイアウト情報登録手段によって登録される文書構造情報が、文書内の各領域の領域名、階層構造、出現順序、繰り返しの有無、及び省略の可否に関する情報を含み、

前記文書抽出手段が、文書内のある箇所が該領域定義情報から得られる該第1のパターンに該当し、かつその箇所に対応する領域名が該文書構造情報から得られる同じ領域名の出現条件を満たす場合に、その箇所を領域の開始箇所と判定し、該第1のパターンと同じかまたは異なるパターンに該当する箇所であって、前記開始箇所とその箇所との間に任意のパターンに該当する箇所がない場合に、その箇所を領域の終了箇所と判定し、該開始箇所と該終了箇所との間の領域を、該文書構造情報から得られる該領域名に対応付けて抽出すること、

を特徴とする請求項1に記載の文書編集システム。

【請求項6】 前記領域定義情報登録手段によって登録される領域定義情報が第1のパターンと第2のパターンから成る1対のパターンを含み、該第1のパターンは前記所望の領域の直前の箇所を特定し、該第2のパターンは所望の領域の直後の箇所を特定し、

前記文書構造・レイアウト情報登録手段によって登録される文書構造情報が、文書内の各領域の領域名、階層構造、出現順序、及び繰り返しの有無、省略の可否に関する情報を含み、

前記文書抽出手段が、文書内のある箇所が該領域定義情報から得られる該第1のパターンに該当し、かつその箇所に対応する領域名が該文書構造情報から得られる同じ領域名の出現条件を満たす場合に、その箇所を前記第1の箇所と判定し、該第1の箇所以降で、該第2のパターンに該当する箇所であって、該第1の箇所とその箇所との間に任意のパターンに該当する箇所がない場合に、その箇所を前記第2の箇所と判定し、該第1の箇所と該第2の箇所との間の領域を、該文書構造情報から得られる該領域名に対応付けて抽出すること、

を特徴とする請求項1に記載の文書編集システム。

【請求項7】 前記領域定義情報登録手段によって登録される領域定義情報が領域の開始の文字列のパターンと領域の終了の文字列のパターンの組を単数または複数含み、

前記文書構造・レイアウト情報登録手段によって登録される文書構造情報が、文書内の各領域の領域名、階層構造、出現順序、繰り返しの有無、及び省略の可否に関する情報を含み、

前記文書抽出手段が、文書内のある箇所が該領域定義情報から得られる該開始の文字列のパターンに該当し、かつその箇所に対応する領域名が該文書構造情報から得られる同じ領域名の出現条件を満たす場合に、その箇所を領域の開始箇所と判定し、該終了の文字列のパターンに該当する箇所であって、前記開始箇所とその箇所との間に任意のパターンに該当する箇所がない場合に、その箇所を領域の終了箇所と判定し、該開始箇所と該終了箇所との間の領域を、該文書構造情報から得られる該領域名に対応付けて抽出すること、

を特徴とする請求項1に記載の文書編集システム。

【請求項8】 前記パターンが、固定的な文字列と、任意の数の任意の文字との組み合わせで定義されること、を特徴とする請求項5ないし7に記載の文書編集システム。

【請求項9】 前記各文書内の領域の構造に関する前記文書構造情報、及び該各領域を再構成するために使用される前記レイアウト情報が、前記文書構造・レイアウト情報登録手段によって一体的に登録されること、

を特徴とする請求項1に記載の文書編集システム。

【請求項10】 少なくとも1つの文書から任意の領域を抽出して、該抽出された領域を編集可能な状態で表示

する文書編集方法において、
各文書から所望の領域を抽出するために使用される領域定義情報をパターンとして登録する領域定義情報登録ステップと、
各文書内の領域の構造に関する文書構造情報、及び該各領域を再構成するために使用されるレイアウト情報を登録する文書構造・レイアウト情報登録ステップと、
該記領域定義登録手段によって登録された領域定義情報と、該文書構造・レイアウト情報登録手段によって登録された文書構造情報を使用して第1の箇所と第2の箇所を検索し、該2つの箇所の間にある領域を抽出する文書抽出ステップと、
該文書構造・レイアウト情報登録手段によって登録された該文書構造情報及び該レイアウト情報を使用して、該抽出された少なくとも1つの領域を再構成して表示する文書再構成ステップと、
を有することを特徴とする文書編集方法。

【請求項11】 少なくとも1つの文書から任意の領域を抽出して、該抽出された領域を編集可能な状態で表示する文書編集方法を実現するプログラムを記録した記録媒体であって、
各文書から所望の領域を抽出するために使用される領域定義情報をパターンとして登録する領域定義情報登録ステップと、
各文書内の領域の構造に関する文書構造情報、及び該各領域を再構成するために使用されるレイアウト情報を登録する文書構造・レイアウト情報登録ステップと、
該記領域定義登録手段によって登録された領域定義情報と、該文書構造・レイアウト情報登録手段によって登録された文書構造情報を使用して第1の箇所と第2の箇所を検索し、該2つの箇所の間にある領域を抽出する文書抽出ステップと、
該文書構造・レイアウト情報登録手段によって登録された該文書構造情報及び該レイアウト情報を使用して、該抽出された少なくとも1つの領域を再構成して表示する文書再構成ステップを、
コンピュータに実行させるプログラムを記録した該コンピュータが読み取り可能な記録媒体。

【請求項12】 前記コンピュータが読み取り可能な記録媒体に記憶された、前記文書編集方法を実現するプログラムが、前記抽出された少なくとも1つの領域が編集された場合、その編集結果を該領域の抽出元の文書に反映させる文書編集結果反映ステップを有すること、
を特徴とする請求項11に記載の記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、文書を編集するシステムに関し、より詳しくは、複数の文書から、同様のパターンに属する領域を抽出し、その抽出された各領域を再構成して表示装置上に編集可能に表示し、更にその

編集後の各領域を抽出元の文書に反映させる文書編集システムに関する。

【0002】

【従来の技術】 従来より、文書を一定の規則で効率よく表示または印刷するための技術としてSGML (Standard Generalized Markup Language) が広く知られている。ユーザはSGMLによって、文書の構造を定義した文書型定義と表示レイアウトまたは印刷レイアウトに使用される文書レイアウト定義を事前に設定することができ、ユーザはこれらの定義を用いて、効率的に一定の規則に従って文書を再構成し、その表示または印刷を行うことができる。前記文書型定義はDTD (Document Type Definition) と呼ばれ、例えば、文書が複数の領域から成る場合、その各領域の階層構造 (包含関係)、出現順及び繰り返し等に関する情報を有している。また前記各領域はそれぞれ任意の一对のタグ (識別子) に関連付けられている。

【0003】 ユーザが文書を作成する場合には、DTDで定義した領域に対応させる部分を、前記領域に関連付けられた一对のタグで挟むように作成する。例えば、文章のタイトル領域は一对のタグ「title」と「/title」に関連付けられており、ユーザは、タイトルとして表示させたい文字列を、この一对のタグで挟むようにする。タグは、場合によっては、対で指定しなくてもよい。

【0004】 前記文書レイアウト定義は、DTDで定義された一对のタグに対応して設定され、例えば、その一对のタグで挟まれた文書内の領域をどういったフォント・スタイルやフォント・サイズで出力するか等を定義する。この文書レイアウト定義は、通常フォーマッタと呼ばれる出力ソフトウェア等によって保持され、フォーマッタが文書処理する際にフォーマッタによって、その文書内に現れる一对のタグに挟まれた各領域が、文書レイアウト定義においてそのタグに関して定義された内容に従って表示または印刷される。例えば、文書内で一对のタグ「title」と「/title」で挟まれた領域は12ポイントのゴシック体で表示される。

【0005】 また、特開平7-98708号公報「文書処理システム及びその方法」には、ユーザが意図した通りに文書またはその文書の一部を選択して、その内容を再構成して表示する文書処理システムが開示されている。この文書処理システムは、ユーザが、例えば「認識技術に役立つ文書を読みたい。」という要求を文書で入力すると、その意図を解析し、その意図に合致した少なくとも1つの文書等 (素材) が選択され、更にその文書の全てまたは一部が再構成されて、ユーザの読みやすい状態に加工された後、表示される。ユーザの意図に合致した文書または文書の一部を選択するために、ユーザが入力した前記文書内の複数のキーワードまたはそのキー

ワードの類義語が用いられる。

【0006】また更に、特開平8-202711号公報「文書編集操作電子装置」には、ユーザが最初に指定した編集対象の範囲から、特定の文字パターンを検出し、同一文書内に同様のパターンを有する部分があれば、その部分も編集対象として自動的に追加登録する装置が開示されている。前記装置は、1つの文書内に散在する見出し、タイトル部分、または注釈文等を一括して編集対象として選択することを目的としている。前記装置は、最初にユーザが編集対象の範囲を1つ選択すると、その範囲内で、見出しなどで多用される「・」や「§」といった特定の記号が存在するかどうかを判定し、それらの記号のうちいずれかが存在する場合は、その記号と記号の表示カラム位置等をパターンとして登録する。次に前記装置は、その文書内で、前記登録されたパターンと同じパターンを有する部分、即ち「・」や「§」といった記号を同じカラム位置に有する部分を検出し、それらを全て編集対象として自動的に選択する。

【0007】

【発明が解決しようとする課題】前述した第1の従来技術SGMLにおいては、文書の一部を抽出し、再構成するために、意識的に一对のタグを文書に埋め込むことが必要であり、このため文書作成の際の労力は多大なものである。この結果、文書内には文書の内容とは直接関係のないタグが散りばめられ、全体的に見映えが悪くなり、場合によってはタグを文書内容と誤解するといった弊害も生じる。また、タグは固定的な文字列であり、これらをパターン化して表示や印刷の際に使用する文書レイアウト定義に対応付けることはできない。

【0008】更に、SGMLでは、1文書を対象とする部分的な抽出と再構成が可能であるが、複数の文書から同様のパターンの文字列に挟まれる領域をそれぞれ抽出して、それらを一覧的に集約して再構成することはできない。また、SGMLは主に、表示または印刷において一定の規則に基づいてレイアウトを行うものであり、再構成された文書の内容を編集可能とし、その編集された内容を元の文書に書き戻すような機能はない。

【0009】また、SGMLでは、文書の構造を定義するDTDと、フォーマット内に保持されたレイアウト定義をそれぞれ設定しなければならない。これはユーザにとって非常に複雑な作業である。更に、タグ名称などはこれらの定義の間で一致させておかななくてはならず、別々に作成することによってタグ名や構造上の矛盾が生じやすくなる。

【0010】前述した第2の従来技術、特開平7-98708号公報「文書処理システム及びその方法」においては、文書の一部を抽出し再構成するために、ユーザの要求として入力された文章から解析されたキーワードまたはそのキーワードの類義語による文字列検索が行われる。こうした文字列検索では、抽出元の文書を少しでも

変更すると、その部分は抽出されなくなってしまう、検索結果が大きく変わってしまう。また、キーワードの類義語も検索しているため、場合によっては検索結果が膨大になり、実用に耐えないケースも生じる。更に、ユーザが要求を文章等で入力するため、所定のパターン化した文字列等を使用して抽出を行うことができない。

【0011】更に、この技術では、複数の文書を対象として抽出を行うことができるが、複数の文書の選択自体がユーザの入力から解析されたキーワード等を用いているので、実際に選択される文書を事前に確定させることができない。また、再構成された文書の内容を編集して、その編集結果を元の文書に書き戻すような機能はない。

【0012】前述した第3の従来技術、特開平8-202711号公報「文書編集操作電子装置」においては、文書の一部を抽出し再構成するために、ユーザが最初に選択した範囲と同じパターンを有する部分が検索されるが、前記パターンは、「・」や「§」といった見出し部分に多用される特定の文字列とその表示カラムであって、汎用的なパターンを用いて検索を行うことはできない。

【0013】更に、この技術では、1文書内における複数の部分を対象にしており、前記パターンを有する文のみが編集対象として選択される。また、編集対象として選択された文に対してレイアウト情報などを用いて再構成するといった機能はない。

【0014】本発明の課題は、汎用的な開始パターン及び終了パターンからなる一对のパターンと、文書の構造を定義する文書構造情報を用いて、複数の文書から前記開始パターンに該当する個所と前記終了パターンに該当する個所との間の領域をそれぞれ抽出し、その抽出結果を前記文書構造情報とレイアウト情報に従って集約して再構成し、ユーザが編集可能な状態で表示する文書編集システムを提供することにある。

【0015】更に、本発明の課題は、編集可能な状態で再構成され表示された領域をユーザが編集した後に、その編集された領域を抽出元の文書に書き戻し、ユーザの編集結果を抽出元の文書に反映させる文書編集システムを提供することにある。

【0016】また更に、本発明の課題は、前記文書構造情報と前記レイアウト情報を一体として登録、保持でき、前記文書構造情報が前記文書内の各領域の領域名、階層構造（包含関係）、出現順序、繰り返しの有無及び省略の可否などの文書構造に関する情報を定義し、前記レイアウト情報が、前記各領域の表示位置、フォント・スタイル、フォント・サイズ、文字列の配置、及び文字色等のレイアウト情報を定義する文書編集システムを提供することにある。

【0017】

【課題を解決するための手段】上記本発明の課題を解決

するための請求項1に記載の文書編集システムは、少なくとも1つの文書から任意の領域を抽出して、該抽出された領域を編集可能な状態で表示する文書編集システムであって、各文書から所望の領域を抽出するために使用される領域定義情報をパターンとして登録する領域定義情報登録手段と、各文書内の領域の構造に関する文書構造情報、及び該各領域を再構成するために使用されるレイアウト情報を登録する文書構造・レイアウト情報登録手段と、該記領域定義登録手段によって登録された領域定義情報と、該文書構造・レイアウト情報登録手段によって登録された文書構造情報を使用して第1の箇所と第2の箇所を検索し、該2つの箇所の間にある領域を抽出する文書抽出手段と、該文書構造・レイアウト情報登録手段によって登録された該文書構造情報及び該レイアウト情報を使用して、該抽出された少なくとも1つの領域を再構成して表示する文書再構成手段を有するよう構成される。これによって、抽出パターンを使用した安定的な領域の抽出が行われ、少なくとも1つの文書から抽出された各領域の内容を集約して見やすい形式で表示することができる。

【0018】請求項2に記載の文書編集システムは、請求項1に記載の前記システムにおいて、前記再構成して表示された少なくとも1つの領域を編集できるように制御する文書編集制御手段を有するよう構成される。これによって、少なくとも1つの文書から抽出された各領域の内容が集約され見やすく表示されると共に、その状態で編集を行うことができる。

【0019】請求項3に記載の文書編集システムは、請求項1に記載の前記システムにおいて、前記文書編集制御手段が更に、前記再構成して表示された少なくとも1つの領域に関する編集機能を、前記文書から抽出された領域の内容のみに制限するよう制御するよう構成される。これによって、抽出された各領域の内容を編集しようとする際に、各領域の内容以外の表示上の修飾部分などを誤って編集することがなく、編集の操作性が向上する。

【0020】請求項4に記載の文書編集システムは、請求項1に記載の前記システムにおいて、前記抽出された少なくとも1つの領域が編集された場合、その編集結果を該領域の抽出元の文書に反映させる文書編集結果反映手段を有するよう構成される。これによって、複数の文書の同様の部分を個別に検索してそれぞれ編集作業を行うことが不要となり、集約的に複数文書の関連部分を編集できる。

【0021】請求項9に記載の文書編集システムは、請求項1に記載の前記システムにおいて、前記各文書内の領域の構造に関する前記文書構造情報、及び該各領域を再構成するために使用される前記レイアウト情報が、前記文書構造・レイアウト情報登録手段によって一体的に登録されるよう構成される。これによって、前記文書

構造情報と前記レイアウト情報が一体として登録、保持され、両者を矛盾なく一体化して容易に作成できるようになった。

【0022】

【発明の実施の形態】以下に、本発明の実施の形態の例について、図面を参照して説明する。尚、各国において、同一または類似のものには同一の参照番号または、記号を付与して説明する。

【0023】図1は、本発明の文書編集システム100の一実施の形態のシステム構成を示すブロック図である。文書編集システム100は、処理装置110、記憶装置120、入力装置130、及び表示装置140を含む。処理装置110は、記憶装置120に記憶されている必要なデータを読み込み、そのデータから得られる情報を元に処理を行い、その処理結果を表示装置140に表示した後、必要に応じてその表示内容を記憶装置120内に記憶された関連するデータに書き込む。

【0024】記憶装置120は、通常ハードディスクやフロッピーディスクなどの2次記憶装置であり、文書編集システム100に必要な、後で詳述する領域定義情報121、文書構造・レイアウト情報122、領域一文書構造対応情報123、及び複数の抽出元となる文書124を格納する。しかし、これらのデータは、ネットワーク経由で他のコンピュータの記憶装置等から読み込まれ、また逆の経路で書き込まれることも可能であり、この場合に前記データが記憶装置120内に格納されている必要はない。

【0025】入力装置130は、必要に応じて記憶装置120に格納されたデータの入力・変更を行うための、マウスやキーボードといった入力装置である。また、入力装置130は、表示装置140上に表示される再構成結果を編集するためにユーザによって使用される。ここで編集とは、表示装置140上に表示された文字列等の対象を入力装置130を使用して追加、変更、または削除することを指す。

【0026】表示装置140は、記憶装置120に格納されたデータや、処理装置110における処理結果の内容を表示するための、CRTディスプレイなどの表示装置である。処理装置110によって処理された再構成結果は最初に、表示装置140に表示される。また、再構成結果やその再構成結果を編集した結果などを印刷するために、印刷装置を設けることも可能である。

【0027】これらの装置の構成は、パーソナル・コンピュータやUNIXワークステーション等において一般的なものである。従って本発明の文書編集システム100も、単一の一般的なパーソナル・コンピュータ等によって実施可能である。

【0028】処理装置110は更に、領域定義情報登録部111、文書構造・レイアウト情報登録部112、文書抽出部113、文書再構成部114、文書編集制御部

115、及び文書編集結果反映部116を含む。

【0029】領域定義情報登録部111は、ユーザが処理対象とする複数の文書のそれぞれから、再構成すべき文書の領域を抽出するための開始パターンと終了パターンからなる1対のパターンを登録する機能を提供する。登録結果は、記憶装置120内の領域定義情報121として記憶される。また、領域定義情報121はテキスト情報でよいから、前記領域定義情報登録部111は、任意のテキストエディタで実現されうるが、所定の入力エリアやチェック機能を備えた特定のアプリケーションによって達成されることが望ましい。

【0030】文書構造・レイアウト情報登録部112は、前記領域定義情報登録部111によって登録された抽出パターンに対応する各領域に対し、領域名、階層構造（包含関係）、出現順序、繰り返しの有無、及び省略の可否などの文書構造情報と、再構成する際の各領域の表示位置、フォント・スタイル、フォント・サイズ、文字列の配置、及び文字色等のレイアウト情報を登録する機能を提供する。登録結果は、記憶装置120内の文書構造・レイアウト情報122として記憶される。この実施の形態では、前記文書構造情報と前記レイアウト情報を一体として登録しているため、登録機構も一体化して示しているが、文書構造情報登録部とレイアウト情報登録部という個別の機構として構成することもできる。またその場合、記憶装置120内の文書構造・レイアウト情報122も、文書構造情報及びレイアウトとして別々に記憶されうる。

【0031】文書抽出部113は、ユーザが指定した記憶装置120内の少なくとも1つの文書124を読み込み、領域定義情報登録部111によって登録された領域定義情報121と文書構造・レイアウト情報登録部112によって登録された文書構造・レイアウト情報122を使用して、文書124の所望の領域を抽出する。文書抽出部113は、読み込んだ前記文書124のそれぞれの中で、前記領域定義情報121に登録された開始パターンと終了パターンからなる1対の抽出パターンに合致する箇所をそれぞれ判定し、更にそれらの箇所に挟まれた領域が、前記文書構造・レイアウト情報122内に定義されている文書内の各領域に関する領域名、階層構造（包含関係）、出現順序、繰り返しの有無、及び省略の可否などの文書構造情報と矛盾していないことを判定すると、その領域を抽出する。また、この抽出の際、抽出元の文書124のどこからどこまでの領域を抽出したかを対応づける情報を、記憶装置120内の領域-文書構造対応情報123に出力する。これは、抽出された各領域が後述する文書編集制御部115によって編集された結果、その領域の長さを変更される可能性があり、その後、これらの領域が後述する文書編集結果反映部116によって個別に抽出元の文書124に反映される際に、文書124内の抽出されていない他の領域を更新してし

まわないようにするためである。

【0032】また、文書抽出部113では、ユーザが文書124を指定した際に、その各文書124の内容をそれぞれ1つのウインドウに表示させ、その後、抽出を指示するよう設定する。しかし、各文書124の内容を表示させることなく、抽出の指定を行うよう構成しても構わない。

【0033】文書再構成部114は、文書抽出部113によって抽出された文書124内の各領域を、領域定義情報登録部111によって登録された領域定義情報121と文書構造・レイアウト情報登録部112によって登録された文書構造・レイアウト情報122を使用して、表示装置140上に再構成する。前記抽出された領域には、それが抽出された抽出パターンに対応する名前（領域名）がつけられており、その領域名に対応する前記文書構造・レイアウト情報122内の内容がそれぞれの領域の再構成に用いられる。また、この再構成に際しては、レイアウト情報によって、抽出された領域に含まれていない見出しなどの固定の文字列や、囲み線、下線等の修飾が付加され、ユーザが見やすいような再構成がなされる。

【0034】また、文書再構成部114は、前記領域-文書構造対応情報123の出力を前記文書抽出部113に代わって行うことも可能である。その場合には、文書抽出部113は、抽出対象となった領域の内容に加えて、それらの領域が抽出元の文書124のどこからどこまでの内容に対応するかを示す中間情報を、文書再構成部114に対して出力しておく必要がある。

【0035】再構成結果は、複数の文書に関して抽出された少なくとも1つの領域が前記レイアウト情報に従って1つのウインドウとして表示される。この実施の形態では、例えばマイクロソフト社のWindows 95のようなマルチウインドウシステム上でワープロWordを稼働させて、その中の1ウインドウに再構成結果を表示させている。また、再構成結果は、1つの文書に関して抽出された少なくとも1つの領域を前記レイアウト情報に従って1つのウインドウとして表示し、それぞれの文書に関して1つのウインドウで表示されるように構成しても良い。

【0036】文書編集制御部115は、前記文書再構成部114によって表示装置140に表示されたウインドウにおいて、そのウインドウ内の編集を制御する。前述の通り、前記ウインドウ内には、複数の文書に関して再構成された結果が表示されているが、抽出された領域と、レイアウト情報によって新たに付加された見出しなどの情報が混在している。そこで、前記文書編集制御部115は、このウインドウ内において、前記抽出された領域のみを編集可能とし、レイアウト情報によって付加された見出しなどの修飾領域の編集を禁止する。また、ここでの編集は、結果的に抽出された領域の長さを変え

てしまうような編集であっても構わない。

【0037】文書編集結果反映部116は、ユーザが抽出元の文書124にその編集結果を反映させるよう指示すると、記憶装置120内の領域—文書構造対応情報123に記憶された対応情報に従って、その抽出元の文書124の内容を編集された内容に更新する。もちろん、反映されるのは、複数の文書から抽出され編集された領域であって、レイアウト情報によって新たに付加された見出しなどの情報は対象外である。

【0038】以上、図1に関して説明を行ってきたが、図示した構成は本発明の一実施の形態の構成を示したに過ぎず、こうした構成に限られるものではない。例えば、処理装置110内の各部分は、複数のコンピュータに分散させて行うことができる。更に、記憶装置120、入力装置130、及び表示装置140は、ネットワークを介して単数または複数の処理装置110と接続させることができ、ユーザがリモートで本発明の文書編集システム100を操作することが可能である。

【0039】次に、図4から図12までを参照しながら図2の処理フローを説明する。図2は、前記文書抽出部113及び前記文書再構成部114による文書抽出及び再構成処理の流れを示す図である。図2は、左1/3の欄が文書抽出及び再構成処理により作成された情報を示し、中央1/3の欄が文書抽出及び再構成処理の流れを示し、右1/3の欄がユーザが予め定義した情報を示している。

【0040】図2の最初のステップS10において、文書抽出部113は領域定義情報121を読み込む。領域定義情報121は、予めユーザが領域定義情報登録部111を用いて登録したものである。

【0041】図4はユーザによって登録された領域定義情報121の一例を示すものである。なお、図面中に記載された黒く塗りつぶされた括弧は本文中では“[]”の括弧に置き換えて表記する。この例では、ある一対の文字列パターン（開始パターンと終了パターン）に該当する文字列に挟まれる文書124内の領域を抽出するように定義されている。例えば、抽出する領域の直前を示す文字列パターン（開始パターン）は【@@】であり、直後を示す文字列パターン（終了パターン）は【@@—終了】であり、【@@】に該当する文字列と【@@—終了】に該当する文字列の間の領域が抽出されるように定義される。 外1 （以下、本文中にお

【0042】

【外1】

【0043】いて、この黒く塗りつぶされた四角は“（黒四角）”と置き換えて表記する。）は任意の領域の終了を示すパターンとして定義されている。但し、ここで、【@@—終了】は必須でなく、【@@】に該当する文字列の次に更に【@@】に該当する別の文字列があ

れば、2つ目の【@@】の直前までが1つ目の【@@】に関する領域の抽出範囲となる。また、@@とは、任意の文字数の文字列を示し、一対の文字列パターンにおける【@@】と【@@—終了】内の@@は同一の文字列である。なお、@@に該当する文字列は、後述する図5の文書構造情報の各領域名（構成要素の名称）と一致するべきものである。

【0044】図2の次のステップS11において、文書抽出部113は文書構造・レイアウト情報122を読み込む。文書構造・レイアウト情報122は、予めユーザが文書構造・レイアウト情報登録部112を用いて登録したものである。ステップS11で使用されるのは、この文書構造・レイアウト情報122のうち、文書124の中の各領域の領域名、階層構造（包含関係）、出現順序、繰り返しの有無及び省略の可否などを定義した文書構造情報のみである。この実施の形態では、文書構造・レイアウト情報122は一体として定義されており、後で図10に関して詳述する。ここでは、説明を簡略化するため、前記文書構造情報の例を概念的に示す図5を用いて説明する。

【0045】図5に示された文書構造情報は、プログラムの仕様書に関する文書構造を規定した例であり、主として文書200の階層構造（包含関係）を示している。最下位の階層のブロック（即ち、見出し、関数名、処理概要、呼出形式、インタフェース、及び注意）が文書内の領域の実体を示している。文書200は、見出し210と処理仕様220から成っている。見出し210は1つ存在する。処理仕様220は繰り返しありと記述されており、複数存在する可能性がある。処理仕様220は、関数名221、処理概要222、呼出形式223、インタフェース224、及び注意225を含んでいるという包含関係を規定する。処理仕様220は前記各要素の集合を規定するのみで、文書200の実体領域に対応しない。インタフェース224は繰り返しありと記述されていることから、1つの処理仕様の中に複数存在する可能性があり、注意225は省略可であり、存在しない場合がある。

【0046】逆に、見出し210、関数名221、処理概要222、呼出形式223、及び少なくとも1つのインタフェース224は、文書200内に必須のものであり、これがない場合は、後述の図2のステップS12において、該当文書に関して所定のタイミングでエラーメッセージを表示したりする対応が考えられる。

【0047】次に図2のステップS12では、文書抽出部113が、ユーザの指定した少なくとも1つの文書124を読み込み、図4の領域定義情報121と図5の文書構造情報に基づいて、各文書124から少なくとも1つの領域を抽出する。前述のように、図5に示された文書構造情報の各領域名（構成要素の名称）が、図4の領域定義情報121のパターンに記載された@@に該当す

る文字列に対応している。即ち文書抽出部 113 は、既に読み取った図 4 の領域定義情報 121 と図 5 の文書構造情報から、ユーザより指定された前記各文書 124 が、少なくとも「見出し」～「見出し-終了」の領域、「関数名」～「関数名-終了」の領域、「処理概要」～「処理概要-終了」の領域、「呼出形式」～「呼出形式-終了」の領域、「インタフェース」～「インタフェース-終了」の領域、及び「注意」～「注意-終了」の領域をこの順序で有し、更に「関数名」～「関数名-終了」の領域から「注意」～「注意-終了」の領域までは、同じ順序で繰り返される可能性があるものとして解釈する。

【0048】但し、前述のように、「見出し-終了」、「関数名-終了」、「処理概要-終了」、「呼出形式-終了」、「インタフェース-終了」、及び「注意-終了」といった領域の終了を判定する文字列は文書 124 中で省略される場合があり、次の領域の開始を判定する文字列または(黒四角)などの所定の文字列の出現により領域の終了が判定される。

【0049】こうした抽出の例を、図 6 に示される文書 124 の例で考えてみる。前述の説明から、文書 124 では、見出し 210 に関する領域は必須となっているが、ここでは省略する。文書抽出部 113 が図 6 の文書 124 を読み込むと、最初に「関数名」という文字列を検出し、その後「関数名-終了」という文字列を検出する。これは、図 4 の領域定義情報 121 のパターンに合致し、図 5 の文書構造情報における関数名 221 の出現順序に合致するため(前述の通り、ここでは見出し 210 はないものと考え、関数名 221 が最初に現れる有効な領域である)、「関数名」～「関数名

-終了」の間の文字列「ZaikoHikiate」が「関数名」という領域名と関連付けて抽出される。

【0050】次に、文書抽出部 113 が読み込みを続けると、「★この行は抽出されません★」という文字列が抽出されるが、この文字列は、図 4 の領域定義情報 121 で定義されたパターンで挟まれておらず、またそのようなパターンで始まっていないため、抽出されない。

【0051】次に、文書抽出部 113 は、「処理概要」という文字列を検出し、その後「呼出形式」という文字列を検出する。ここで、「処理概要」という文字列は図 4 の領域定義情報 121 のパターンに合致し、図 5 の文書構造情報において、関数名 221 の次に出現する領域名が処理概要 222 であることから出現順序の条件も満たし、「処理概要」～「呼出形式」の間の文字列「(改行) 与えられた商品コードと受注数量を基に、在庫が引き当て可能かどうか(改行)を判断する。」が、「処理概要」という領域名と関連付けて抽出される。この場合、「処理概要-終了」という、領域の終了を示す文字列はないが、次の妥当な領域「呼出形式」の開始を示す「呼出形式」という文字列が出現したため前記文字列「処理概要-終了」が省略されているものと判断される。また、ここで抽出される文字列の内容には、抽出元の内容を忠実に再現すべく、文字情報の他に改行や改ページに関する制御情報も抽出される。前記処理概要の例では、改行に関する制御情報が、抽出文字列内に「(改行)」として抽出されている。

【0052】以降、同様の処理を繰り返すと、その後、以下のような文字列が各領域名に関連付けて抽出される。

```
領域名：呼出形式＝「ZaikoHikiate (ShohinCD, ZaikoSu, Kekka)」、
領域名：インタフェース (1)＝「Int ShohinCD, //
引き当てすべき商品コードのキーとなる値」、
領域名：インタフェース (2)＝「Int ZaikoSu, //
引き当てすべき数量」、
領域名：インタフェース (3)＝「Int Kekka //
引き当てが可能かどうかセットされる。(改行) //
引き当て可能なら 1。不可なら 0 が設定される。」、
領域名：注意＝「在庫数 (ZaikoSu) にマイナスの数値が入っている場合、本処理の動作結果(改行)は不定になる。」。
```

【0053】ここで、処理仕様 220 の繰り返しはないがインタフェース 224 は 3 回繰り返されており、このような場合、領域名には、前述のようにインタフェースの後に (1)～(3) が付加される。これは、繰り返しの要素を固有に識別可能にするためであって、他の様々な方法を用いることもできる。

【0054】図 7 には、領域定義情報 121 の第 2 の例が示されている。この例では、ある程度規則的な文字列のパターンを有する文書 124 内の文字列から開始され

る領域が抽出されるように定義されている。例えば、抽出する領域の直前の文字列は「[0-9]+¥, @@ \$」であり、領域の直後の文字列を示すパターンはない。「[0-9]+¥, @@ \$」は、行頭から少なくとも 1 つの数字の並び、1 つのピリオド、領域名に対応する任意の数の文字列、及び改行からなる文字列を示している。任意の領域の終了を示すパターンは、外 2 である。ここで、^ \$ は改

【0055】

【外2】

^ \$ ■

【0056】行のみの行を示している。従って、改行のみの行または文字列（黒四角）が現れた場合に任意の領域が終了する。前述のように@@は、文書構造情報の各領域名（構成要素の名称）と同じものである。

【0057】図7の領域定義情報121に基づいた抽出の例を、図8に示す第2の文書124の例で考えてみる。また、ここでは、図5に示したような文書構造情報が省略されているが、階層構造、及び繰り返しを持たない「発明の名称」、「特許請求の範囲」、及び「発明の詳細な説明」の各領域が、この順で並んでいる単純な文書構造を仮定する。

【0058】文書抽出部113が図8の文書124を読み込むと、最初に「1. 発明の名称」という文字列を検出し、その後改行のみの行を検出する。ここで、図7の領域定義情報121から、@@に対応するのは「発明の名称」であり、前記仮定した文書構造情報において文書内で最初にあるべき領域「発明の名称」に合致するため、「1. 発明の名称」～改行のみの行の間にある文字列「（改行）電子ファイル編集装置（改行）」を「発明の名称」という領域名に関連付けて抽出する。以降同様に抽出を繰り返すと、その後、以下のような文字列が各領域名に関連付けて抽出される。

【0059】領域名：特許請求の範囲＝「（改行）電子ファイルの領域を指定するためのルールと、その領域をレイアウトする情報を定義することにより、電子ファイルを再構成し、編集する装置（改行）」、
領域名：発明の詳細な説明＝「（改行）本発明は、複数...（改行）は不足になる。」。

【0060】図9には、領域定義情報121の第3の例が示されている。この例では、ある程度固定的な文字列のパターンを有する文書124内の文字列に挟まれた領域が抽出されるように定義されている。例えば、「1. 発明の名称」という文字列と、^ \$（改行のみの行）、または（黒四角）との間の文字列は、発明名称という領域名で抽出され、「2. 特許請求の範囲」という文字列と、^ \$または（黒四角）との間の文字列は、請求範囲という領域名で抽出され、「3. 発明の詳細な説明」という文字列と、^ \$または（黒四角）との間の文字列は、詳細説明という領域名で抽出される。ここで、^ \$は改行のみの行を示している。

【0061】図4または図7に示した領域定義情報の例では、文書中のパターンの中の一部（@@に対応する部分）が領域名に対応していたが、この第3の例では、パターン毎に固定的に設定されている。また、この例でも、文書構造情報が省略されているが、階層構造、及び繰り返しを持たない「発明名称」、「請求範囲」、及び「詳細説明」の各領域が、この順で並んでいる単純な文書構造を仮定する。ここでは、各領域の名称が前記領域

定義情報121内の各抽出パターンに対応して設定されており、図8の説明で仮定した文書構造情報における領域名とは異なっていることに注意すべきである。

【0062】図9の領域定義情報121を用いて、再び図8の文書124からの抽出を行うと、文書抽出部113は最初に「1. 発明の名称」という文字列を検出し、その後改行のみの行を検出する。ここで、図9の領域定義情報121から、対応する領域名は「発明名称」であり、前記仮定した文書構造情報において文書内で最初にあるべき領域「発明名称」に合致するため、「1. 発明の名称」～改行のみの行までの文字列「（改行）電子ファイル編集装置（改行）」を「発明名称」という領域名に関連付けて抽出する。以降同様に抽出を繰り返すと、その後、以下のような文字列が各領域名に関連付けて抽出される。

【0063】領域名：請求範囲＝「（改行）電子ファイルの領域を指定するためのルールと、その領域をレイアウトする情報を定義することにより、電子ファイルを再構成し、編集する装置（改行）」、
領域名：詳細説明＝「（改行）本発明は、複数...（改行）は不足になる。」。

【0064】図2のステップS12で、文書抽出部113が文書124から抽出対象の領域を抽出する方法を以上に示したが、文書抽出部113は、この抽出と共に、抽出元の文書124のどこからどこまでの領域を抽出したかを対応づける情報を、記憶装置120内の領域一文書構造対応情報123に出力する。図10は、前記領域一文書構造対応情報123の内容の一例を示す図である。

【0065】領域一文書構造対応情報123は、前記文書構造情報における領域名毎に、その領域が抽出元の文書124のどの位置に対応するかを示す抽出元文書情報を有している。ここで、図5に示す文書構造情報を有する文書124を仮定すると、「見出し」に対応する抽出元文書情報は、「" C: ¥文書 ¥ source-1. o", From (2, 10), To (2, 20)」であり、これは、抽出元の文書が、記憶装置120内のODライブ内の「文書」というディレクトリ内の「source-1. o」というファイルであり、抽出された「見出し」の領域をその文書の2行目の10カラム～2行目の20カラムの位置から抽出したということを示している。他の領域も同様であるが、図10では、処理仕様及びインタフェースに繰り返しがあり、領域名においては、それぞれ繰り返し要素が括弧内に表されている。また、その繰り返し要素と階層関係（包含関係）を表すために、前記領域名は、例えば「処理仕様（2）、インタフェース（1）」といった連結された表記となっている。

【0066】この例では、抽出された領域の抽出元の文書124における位置を、各領域の開始行とカラム、及

び終了行とカラムを保持することによって記憶するようにしているが、各領域名、即ち検索パターン毎の抽出順（この場合は、書き戻す際に、文書124に対して前記文書抽出部113による抽出処理と同様のパターン検索処理が必要）や、抽出元の文書124における開始アドレス位置と抽出された領域のサイズを保持することによって、前記抽出された領域の抽出元の文書124における位置を記憶してもよい。

【0067】こうして、抽出された各領域の抽出元の文書124における位置が記憶されることによって、即ち、領域-文書構造対応情報123を保持することによって、前記文書抽出部113によって抽出された領域が、抽出元の文書124と独立して編集された場合でも、文書124内の抽出されていない他の領域との整合性を保ちつつ、前記抽出され編集された領域を文書124に反映することが可能となる。

【0068】再び図2のステップS13に戻ると、ここでは前記文書再構成部114が、前記文書抽出部113が文書124から抽出した領域の情報を読み込み、文書構造・レイアウト情報122に基づいて、表示装置140上に、それらの領域の情報を再構成して表示する。

【0069】図11に文書構造・レイアウト情報122の例を示す。この文書構造・レイアウト情報122は、図5に示した概念的な文書構造情報と、前記再構成のために使用されるレイアウト情報を同時に定義するための方法を示したものである。前述したように、文書構造情報とレイアウト情報は、個別に定義されていても良い。

【0070】この例では、マイクロソフト社のワープロWordを使用して、各領域に関して、その領域の表示位置、フォント・スタイル、フォント・サイズ、文字列の配置、及び文字色等のレイアウト情報と、領域名、包含関係、繰り返しの有無、及び省略の可否を示す文書構造情報を定義し、その他囲み線や区切り線などのワープロによる修飾を付加している。これらの情報のうち、フォント・スタイル、フォント・サイズ、文字列の配置、及び文字色等のレイアウト情報と、領域名、繰り返しの有無、及び省略の可否を示す文書構造情報は、各領域毎にWordのブックマークの機能を利用して設定している。

【0071】しかし、このように、各領域に対して、レイアウト情報及び文書構造情報を定義可能であり、その情報をコンピュータによって制御（入出力を含む）できれば、他のどのような手段を用いても構わない。また、マイクロソフト社のWordを用いた本発明の例は、単に例示のためのものに過ぎず、こうした具体的な手段のみに制限されるものではない。

【0072】図11に示す文書構造・レイアウト情報122は、まず可視部分と不可視部分に大別される。可視部分は更に、ユーザによる編集が禁止される固定部分と、抽出された領域が表示される、ユーザによる編集が

可能な編集部分に分類される。区切り線326、固定タイトル327、328、329、及び32A、及び各領域を囲む囲み線は固定部分である。領域310、321、322、323、324、及び325は、それぞれ見出し、関数名、処理概要、呼出形式、インタフェース、及び注意の各領域名称に対応する編集部分である。

【0073】前述のように、文書抽出部113は、前記文書124から抽出した領域をそれぞれ見出しや関数名といった領域名と関連付けて、抽出された領域の情報として出力する。前記文書再構成部114は、文書抽出部113から受け取った抽出された領域の情報を、対応する領域名を有する図11に示された領域の位置に出力する。この領域は、ユーザによる編集が可能で、前記編集制御部115によって、表示された内容の削除、変更、追加を行うことができる。前記編集制御部115は、その他の領域のユーザによる編集を禁止する。所定の領域のみ編集を有効にすることは、従来のワープロの機能を使用して容易に達成することができるが、特別なアプリケーションを用意して、ワープロ上に組み込むこともできるし、全く独立した別のアプリケーションとして作成することもできる。

【0074】不可視部分には、上位階層の領域である処理仕様に対応する領域320と、各領域に関してレイアウト情報、及び文書構造情報を定義する定義部分330、331、332、333、334、335、及び336がある（この定義内容は、Wordのブックマーク機能によって設定されている）。領域320は、図5の文書構造情報に示す処理仕様220に対応するものであり、前記編集領域321、322、323、324、及び325の集合を含むように示されている。このことによって、前記各領域320及び321～325の階層構造（包含関係）が分かり、これも1つの文書構造情報である。この領域320の定義をしているのが定義部分336である。ここでは、この領域320が処理仕様という領域名称を有し、繰り返し有りであることが定義されている。領域320が、各領域321から325を包含しているということは、各領域321から325を全て含むように領域320が設定されていることで示されている。

【0075】その他の定義部分330、331、332、333、334、及び335は、それぞれ編集領域310、321、322、323、324、及び325に関するものであり、表示される編集対象の文字列に関するレイアウト情報及び文書構造情報を含んでいる。

【0076】例えば、定義部分330領域310に対応し、表示される文字列のフォント・スタイルは明朝体、フォント・サイズは22ポイント、文字修飾はBOLD（太字）とし、文字列の配置は中央揃え（センタリング）、色は黒で、対応する領域310の領域名称が「見出し」であることを示している。

【0077】また、領域324に対応する定義部分334は、領域324の領域名称がインタフェースであり、繰り返し有りであることを示している。領域325に対応する定義部分335は、領域326の領域名称が注意であり、省略可であることを示している。

【0078】文書再構成部114が、このようにして文書抽出部113から抽出された領域の情報を再構成して表示画面140上に表示した表示例400を図12に示す。図12の表示例400は、図6に示した文書124を、図4に示す領域定義情報121、及び図11に示す文書情報・レイアウト情報122に基づいて抽出し、再構成し、表示したものである(但し、図6の文書124では、見出し「Aプロジェクト-関数定義仕様書」は省略してある)。

【0079】この表示例400では、1つの文書から複数の領域が抽出された場合を示している。複数の文書から領域が抽出された場合は、この表示の下に追加されるか、または別のウィンドウ内の表示として提供される。このことによって、ユーザは、例えば、複数のプログラムの仕様書の処理概要や呼出形式等の内容を同じ様式で一覧表示させる事ができる。

【0080】図3には、文書編集結果反映部116による文書編集結果反映処理の流れが示されている。図3は、左1/3の欄が再構成処理及びユーザの編集により作成された情報を示し、中央1/3の欄が文書編集結果反映処理の流れを示し、右1/3の欄がユーザが予め定義した情報を示している。

【0081】ユーザが編集結果を抽出元の文書124に反映させるよう指示すると、図3の最初のステップS20において、文書編集結果反映部116が文書構造情報・レイアウト情報122を読み込む。これは、各編集領域の領域名等を取得するためであり、この実施の形態では、前述のように文書構造情報・レイアウト情報122の保存及び編集にWordの機能が使用されているため、Wordの関係ファイルまたはメモリから読み込まれる。

【0082】次に、ステップS21において、文書編集結果反映部116は、文書抽出部113によって出力された領域-文書構造対応情報123を読み込み、ユーザによって編集された表示装置140上の各領域の内容を、抽出元の文書124のどの位置に反映させるかを判定する。

【0083】次に、ステップS22において、文書編集結果反映部116は、ユーザによって編集された表示装置140上の各領域の内容を、抽出元の文書124の対応する位置に反映させる。ユーザが表示装置140上のある領域に対して任意の文字数だけ削除を行った場合は、その領域を抽出元の文書124に反映させた時にその領域の最終位置となるアドレスAが、領域-文書構造対応情報123に保持されているその領域の抽出時に

ける最終位置のアドレスBより小さいので、その領域を抽出元の文書124に反映させた後、両者のアドレスの差だけ、抽出元の文書124のアドレスB以降の内容をアドレスAの後に移動させる必要がある。

【0084】ユーザが表示装置140上のある領域に対して任意の文字数だけ追加を行った場合は、その領域を抽出元の文書124に反映させた時にその領域の最終位置となるアドレスAが、領域-文書構造対応情報123に保持されているその領域の抽出時における最終位置のアドレスBより大きいので、その領域を抽出元の文書124に反映させる前に、両者のアドレスの差だけ、抽出元の文書124のアドレスB以降の内容をアドレスAの後に移動させておく必要がある。

【0085】ユーザが表示装置140上のある領域に対して編集を行わなかったかまたは、編集を行ったものの結果的に文字数に変動がなかった場合は、その領域を抽出元の文書124に反映させた時にその領域の最終位置となるアドレスAと、領域-文書構造対応情報123に保持されているその領域の抽出時における最終位置のアドレスBは等しいので、その領域を抽出元の文書124の元の位置に反映させるだけでよい。

【0086】前記アドレスAやアドレスBは、図10に例示した領域-文書構造対応情報123に示すように、文書124における行とカラムで把握することができるが、文書124の先頭からの論理アドレスとして把握する事も可能である。

【0087】次に、図13から図20を参照して、本発明の文書編集システムの一連の操作をより具体的に説明する。また、領域定義情報121は図4に示したものの、文書構造・レイアウト情報122は図11に示したものが使用される。

【0088】図13には、抽出元の文書124の読み込みを指定する画面410が示されている。このように、本発明の文書編集システムは、例えばマイクロソフト社のワープロWordなどに本発明の独自の機能を組み込むことによって達成することができるが、新たに作成してもよく、また任意の方法で既存のアプリケーションと組み合わせて実現することも可能である。

【0089】画面の表示例410のメニューで「表示(V)」をマウスのクリックやキーボード等によって選択すると、図示するようなサブメニュー411が表示される。そこで、このサブメニュー411から「抽出ファイル指定(S)」を選択すると、図14に示すような抽出元の文書124を選択する画面421が表示される。

【0090】図14は、抽出元の文書124を選択する画面421を示している。画面421は画面420からのポップアップ・ウィンドウとして表示される。この例では、Title.txt(422)、ZaikoHiklate.c(423)、及びZaikoKoushin.c(424)の3ファイルが選択されている。これ

らのファイルはいずれもテキスト・ファイルである。本発明が抽出元の文書124に関して文字列のパターン検索を行い、抽出結果の編集を可能にするシステムであることから、ここでテキスト情報を有しないバイナリ・ファイルなどを指定することに意味はない。この画面421で抽出の対象となる抽出元の文書124を選択した後「開く(O)」ボタンを選択すると、図15に示すように、選択された抽出元の文書124の内容を表示する画面430が表示される。

【0091】図15は、選択された抽出元の文書124の内容を表示する画面430を示している。指定された文書124の内容がそれぞれ1つのウインドウ内に表示されている。ウインドウ431には、ZaikoKoushin, cの内容が、ウインドウ432には、ZaikoHikiate, cの内容が、ウインドウ433にはTitle, txtの内容がそれぞれ表示されている。

【0092】図16は、図15の状態から再構成処理を指示する画面440を示している。画面440のメニューで「表示(V)」を選択すると、図示するようなサブメニュー444が表示される。そこで、このサブメニュー444から「ソース→仕様(S)」を選択すると、選択された全ての文書124に関して再構成処理が開始される。設計によっては、図15の内容表示画面を省略することも可能である。

【0093】図17は、図16で再構成の指定がされた場合に、再構成結果を表示する画面450を示している。画面450の再構成結果は1ウインドウとして表示され、抽出元の文書124の区切りはページ区切りで表される。例えば、ZaikoHikiate, cから抽出され再構成された内容は、領域451に示され、ZaikoKoushin, cから抽出され再構成された内容は、領域452に示されている。このように、抽出元の文書124の区切りをページ区切りによって表すのに代えて、抽出元の文書124毎に別のウインドウで表示するような態様も可能である。この画面450においては、図11に示す領域310、321、322、323、324、及び325に対応する部分に関して、ユーザはワープロと同様の方法で自由に編集作業を行うことができ、それ以外の領域にはカーソルが移動せず、ユーザによる編集ができないように制御されている。

【0094】図18は、図17で示された再構成結果の表示画面450において、ZaikoHikiate, cから抽出され再構成された内容を編集している画面460を示している。ここでの編集は、処理概要に関する記述内容を付加するものである。ユーザの編集によって加えられた部分461は、それが再構成された後に編集されたものであることを明示するために、この例のように反転表示されているのが望ましい。

【0095】図19は、編集結果を抽出元の文書124に反映させる指示を行う画面470を示している。画面

470のメニューで「表示(V)」を選択すると、図示するようなサブメニュー472が表示される。そこで、このサブメニュー472から「仕様→ソース(D)」を選択すると、元の文書124に対してその編集結果が反映される。この例では、抽出元の文書ZaikoHikiate, cに対して、実質的に内容が追加された部分(反転領域471)が反映される。

【0096】図20は、編集結果が反映された抽出元の文書124の内容を示す画面480を示している。図19に示す画面470の操作で編集結果が反映された文書ZaikoHikiate, cを再び表示させると、ウインドウ481に示すように、処理概要の内容に、「その際...チェックする。」という文が追加されている。また、図3のステップS22において、文書編集結果反映部116が、ユーザによる編集結果を抽出元の文書124に反映させた時に、その編集された部分が最終的に抽出元の文書124のどの位置に反映されたかを記憶しておけば、ウインドウ481に示すように、ユーザによって追加等された領域を反転表示させることも可能である。

【0097】図21は、上述した文書編集システム100を構築するために使用されるコンピュータ600のハードウェア構成の一例を示している。該コンピュータ600は、それぞれバス680に接続されたCPU610、記憶部620、メモリ部630、表示部640、入力部650、印刷部660、及びネットワーク・インタフェース部670からなる。

【0098】CPU610は、図1の文書編集システム100の処理装置110に対応し、領域定義情報登録部111、文書構造・レイアウト情報登録部112、文書抽出部113、文書再構成部114、文書編集制御部115、及び文書編集結果反映部116の各部分を実行する。

【0099】記憶部620は、図1の文書編集システム100の記憶装置120に対応し、CPU610によって実行される前記各機能を実現するプログラム、及び領域定義情報121、文書構造・レイアウト情報122、領域-文書構造対応情報123、及び複数の抽出元となる文書124を格納するメモリ部630には、CPU610によって実行される前記各部分を実現するプログラムがロードされ、また必要に応じてユーザの編集内容などを含む表示部640に表示されている内容や、記憶部620内の各種情報の内容がロードされる。

【0100】表示部640は、図1の文書編集システム100の表示装置140に対応し、再構成結果を表示し、それをユーザが編集する際に必要とされ、通常はCRTやLCD等のディスプレイ装置である。

【0101】入力部650は、図1の文書編集システム100の入力装置130に対応し、前記表示部640に表示された画面に沿って入力や指示を行うために使用さ

れる装置であり、通常キーボード、マウス等から成る入力装置やタッチパネル、音声入力装置等で構成される。

【0102】印刷部660は、ユーザ等の指示に従って、前記記憶部620またはメモリ部630に格納されているデータ等を印刷する、レーザプリンタ等の印刷装置である。前記印刷部660は、前記文書編集システム100の実施に関して必須の構成要件ではない。

【0103】ネットワーク・インタフェース部670は、主に、リモートにある記憶部620との接続、または他のCPUとの接続を実現するよう機能する。前記文書編集システム100を単一のコンピュータで実施する場合には必要とされない。

【0104】バス680は、前記各構成要素610～670間でデータ、指令等の送受信を行うための共通伝送経路である。

【0105】

【発明の効果】本発明の文書編集システムによれば、汎用的な開始パターン及び終了パターンからなる一対のパターンと、文書の構造を定義する文書構造情報を用いて、複数の文書から前記開始パターンに該当する個所と終了パターンに該当する個所との間の領域がそれぞれ抽出され、その抽出結果が、前記文書構造情報とレイアウト情報に従って集約して再構成され、ユーザが編集可能な状態で表示される。この結果、複数の文書の情報を、ユーザが最も見やすいレイアウトで表示させたまま編集作業を行うことができ、複数の文書から必要な領域を検索して直接編集作業を行うことに比べて格段に編集作業が効率化された。

【0106】また、抽出対象を文字列の一致などではなく、抽出領域の開始、終了を示す文字列パターンを用いて指定することにより、複数の文を含む広い領域を指定することができる。文字列の一致などの方法では、抽出元の文書を少し変えただけでも抽出結果が大きく変化するのに対し、本発明のシステムでは開始や終了パターンに該当する部分に変更されない限り、抽出される領域自体に変化はない。この結果、常にユーザの意図した抽出結果を得ることができ、安定した再構成結果を表示させることができる。

【0107】更に本発明によって、編集可能な状態で再構成され表示された領域がユーザに編集された後、その編集された領域が抽出元の文書に、その文書の他の領域と矛盾することなく書き戻される。

【0108】また更に本発明によって、前記文書構造情報と前記レイアウト情報が一体として登録、保持され、前記文書構造情報には前記文書内の各領域の領域名、階層構造（包含関係）、出現順序、繰り返しの有無、及び省略の可否などの文書構造に関する情報が含まれ、前記レイアウト情報には、前記各領域の表示位置、フォント・スタイル、フォント・サイズ、文字列の配置、及び文字色等のレイアウト情報が含まれる。従来は両方の前記

情報を突き合わせながら、両者が矛盾しないように注意深く作成していたが、本発明により、両者を矛盾なく一体化して容易に作成できるようになった。

【図面の簡単な説明】

【図1】本発明の文書編集システムの一実施の形態のシステム構成を示すブロック図である。

【図2】文書抽出及び再構成処理の流れを示す図である。

【図3】文書編集結果反映処理の流れを示す図である。

【図4】領域定義情報の例を示す図である。

【図5】文書構造情報の例を概念的に示す図である。

【図6】抽出元の文書の例を示す図である。

【図7】領域定義情報の第2の例を示す図である。

【図8】抽出元の文書の第2の例を示す図である。

【図9】領域定義情報の第3の例を示す図である。

【図10】領域一文書構造対応情報の例を示す図である。

【図11】文書構造・レイアウト情報の例を示す図である。

【図12】再構成結果の表示例を示す図である。

【図13】抽出元の文書の読み込みを指定する画面の例を示す図である。

【図14】抽出元の文書を選択する画面の例を示す図である。

【図15】選択された抽出元の文書の内容を表示する画面の例を示す図である。

【図16】再構成処理を指示する画面の例を示す図である。

【図17】再構成結果を表示する画面の例を示す図である。

【図18】表示された再構成結果に対して編集を行った画面の例を示す図である。

【図19】編集結果を抽出元の文書に出力するよう指示する画面の例を示す図である。

【図20】編集結果が反映された抽出元の文書の内容を示す画面の例を示す図である。

【図21】文書編集システムを実行するコンピュータのハードウェア構成を示す図である。

【符号の説明】

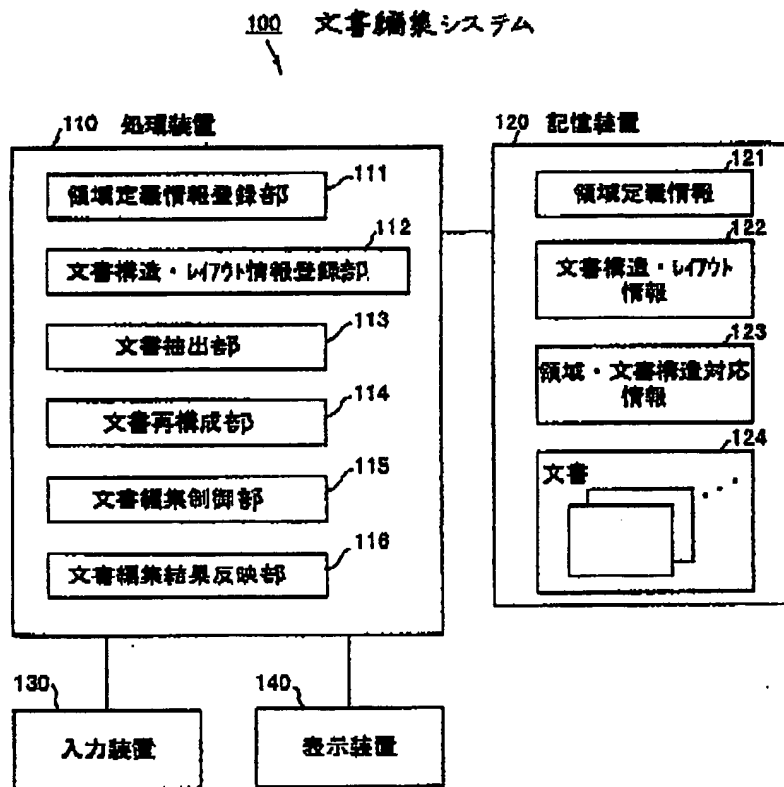
- 100 文書編集システム
- 110 処理装置
- 111 領域定義情報登録部
- 112 文書構造・レイアウト情報登録部
- 113 文書抽出部
- 114 文書再構成部
- 115 文書編集制御部
- 116 文書編集結果反映部
- 120 記憶装置
- 121 領域定義情報
- 122 文書構造・レイアウト情報

123 領域—文書構造対応情報
124 文書

130 入力装置
140 表示装置

【図1】

本発明の文書編集システムの
一実施の形態のシステム構成を示すブロック図



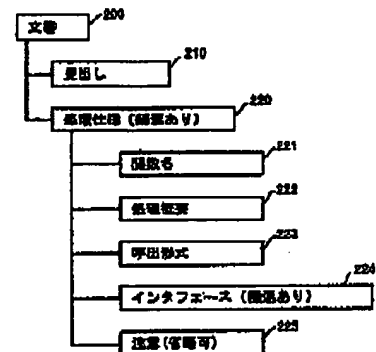
【図4】

領域定義情報の例を示す図

121 領域定義情報	パターン
パターンの説明	
文書の領域の開始を示す文字列パターン	【@@】
文書の領域の終了を示す文字列パターン	【@@-終了】
任意の領域の終了を示す文字列パターン	■

【図5】

文書構造情報の例を概念的に示す図



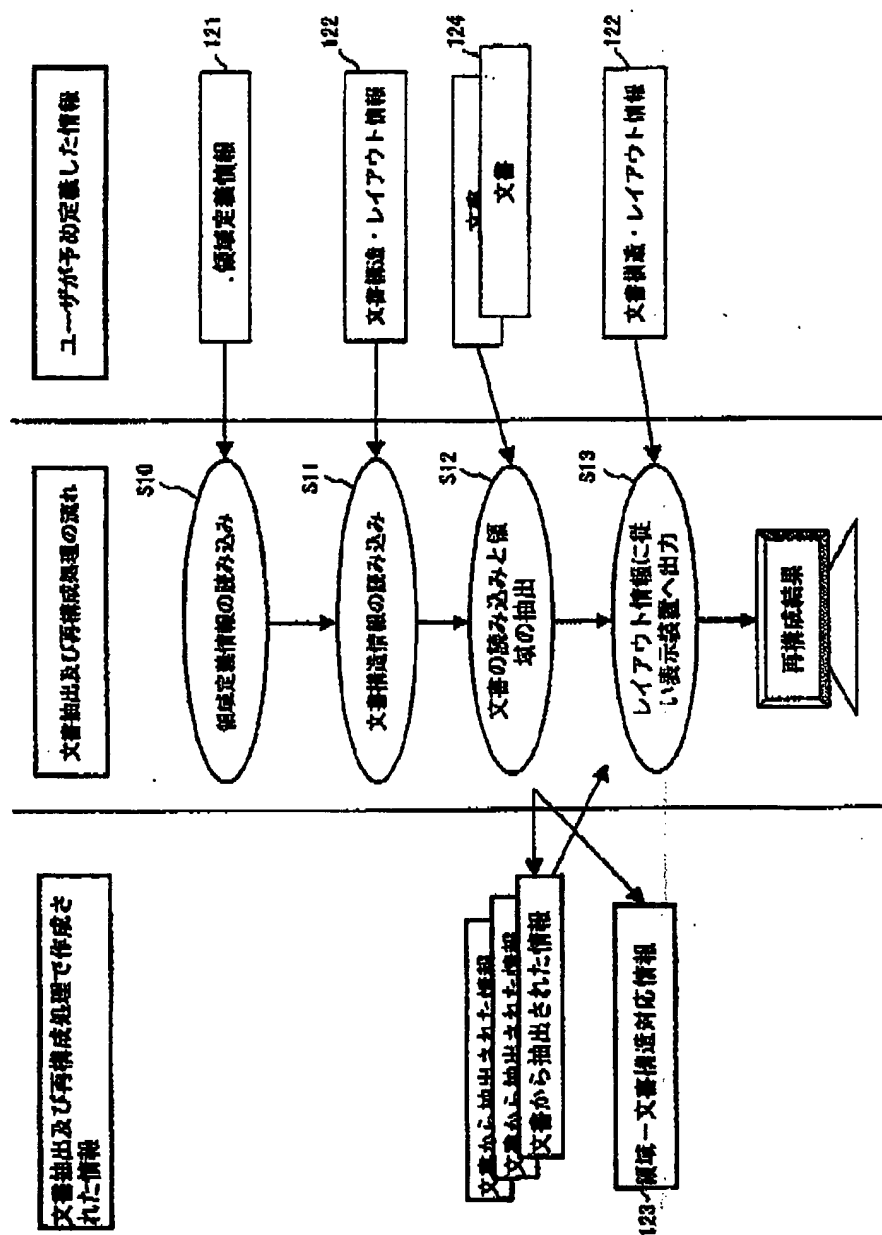
【図7】

領域定義情報の第2の例を示す図

121 領域定義情報	パターン
パターンの説明	
文書の領域の開始を示す文字列パターン	*[0-9]+, @@
文書の領域の終了を示す文字列パターン	なし
任意の領域の終了を示す文字列パターン	* 省略

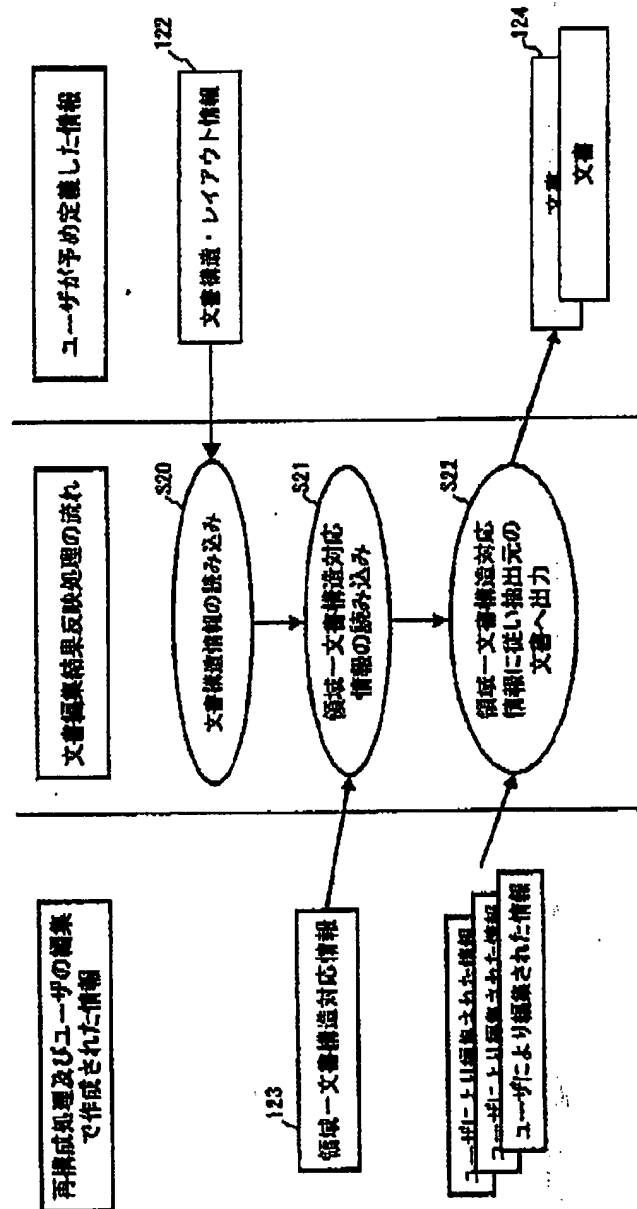
【図 2】

文書抽出及び再構成処理の流れを示す図



【図3】

文書編集結果反映処理の流れを示す図



【図 6】

文書の例を示す図

文書 124

【関数名】 ZaikoHikiato 【関数名一桁7】

★この行は抽出されません。★

【処理概要】

与えられた商品コードと受注数量を基に、在庫が引き当て可能かどうかを判断する。

【呼出形式】

ZaikoHikiato (ShohinCD, ZaikoSu, Kakka)

【インターフェース】

int ShohinCD, //引け当てすべき商品コードのキーとなる値
int ZaikoSu, //引け当てすべき数量
int Kakka, //引け当てが可能かどうかセットされる。
//引け当て可能なら1、不可なら0が返される。

【注意】

在庫数 (ZaikoSu) にマイナスの数値が入っている場合、本処理の動作結果は不定になる。■

【図 9】

領域定義情報の第3の例を示す図

121. 領域定義情報

領域名称	開始パターン	終了パターン
処理概要	1. 処理概要	★ 終
呼出形式	2. 呼出形式	★ 終
処理概要	3. 処理概要	★ 終

【図 8】

抽出元文書の第2の例を示す図

124 文書

1. 処理概要
電子ファイル検索結果
2. 呼出形式の範囲
電子ファイルの領域を指定するためのルールと、その領域をレイアウトする情報も記載することにより、電子ファイルを再現し、読取する機能
3. 処理概要の説明
本処理は、検索...
は不定になる。■

【図 10】

領域一文書抽出対応情報の例を示す図

123 領域一文書抽出対応情報

領域名	抽出元文書情報
抽出し	"0: 文書名Source-1.0", From(1.10), To(2.20)
抽出仕組 (1), 抽出概要	"0: 文書名Source-1.0", From(1.1), To(2.20)
抽出仕組 (1), 抽出概要	"0: 文書名Source-1.0", From(1.1), To(2.20)
抽出仕組 (1), 呼出形式	"0: 文書名Source-1.0", From(1.1), To(2.20)
抽出仕組 (1), インタフェース (1)	"0: 文書名Source-1.0", From(1.1), To(2.20)
抽出仕組 (1), インタフェース (2)	"0: 文書名Source-1.0", From(1.1), To(2.20)
抽出仕組 (1), インタフェース (3)	"0: 文書名Source-1.0", From(1.1), To(2.20)
抽出仕組 (2), 抽出概要	"0: 文書名Source-1.0", From(1.10), To(2.20)
抽出仕組 (2), 抽出概要	"0: 文書名Source-1.0", From(1.1), To(2.20)
抽出仕組 (2), 呼出形式	"0: 文書名Source-1.0", From(1.1), To(2.20)
抽出仕組 (2), インタフェース (1)	"0: 文書名Source-1.0", From(1.1), To(2.20)

【図 12】

再現結果の表示例を示す図

400

Aプロジェクトー関数定義仕様書

ZaikoHikiato

●処理概要

与えられた商品コードと受注数量を基に、在庫が引き当て可能かどうかを判断する。

●呼出形式

ZaikoHikiato (ShohinCD, ZaikoSu, Kakka)

●インターフェース説明:

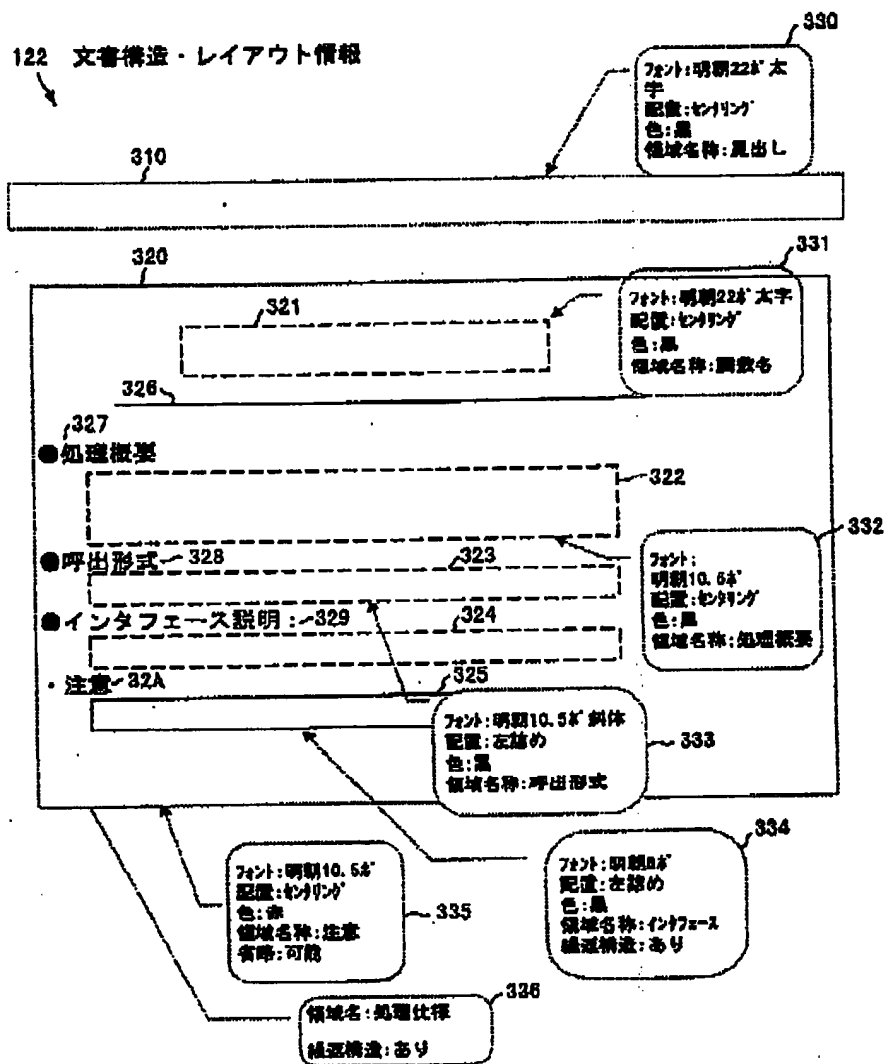
int ShohinCD, //引け当てすべき商品コードのキーとなる値
int ZaikoSu, //引け当てすべき数量
int Kakka, //引け当てが可能かどうかセットされる。
//引け当て可能なら1、不可なら0が返される。

●注意

在庫数 (ZaikoSu) にマイナスの数値が入っている場合、本処理の動作結果は不定になる。

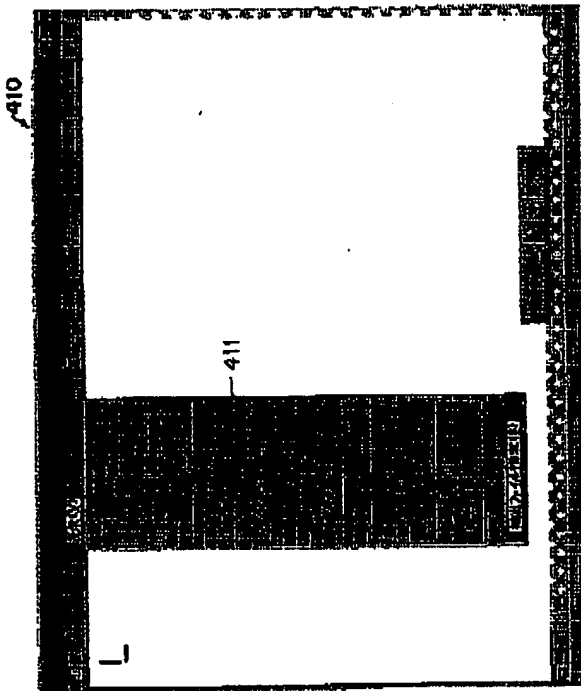
【図11】

文書構造・レイアウト情報の例を示す図



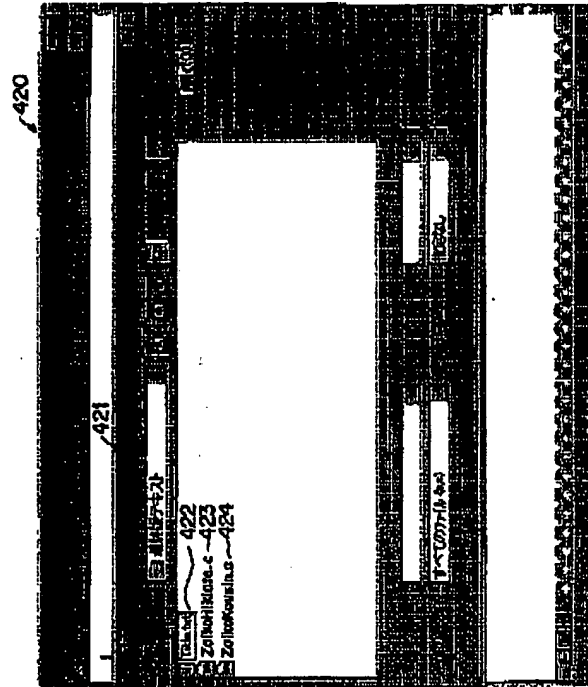
【図13】

抽出元の文書の読み込みを指定する
画面の例を示す図



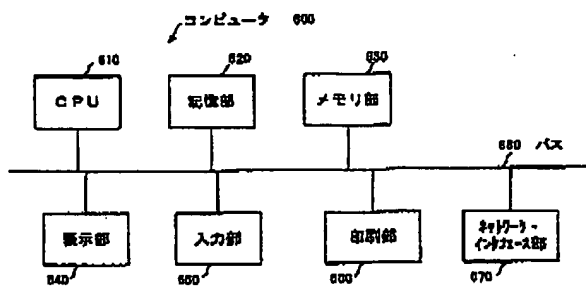
【図14】

抽出元の文書を選択する画面の流れを示す図



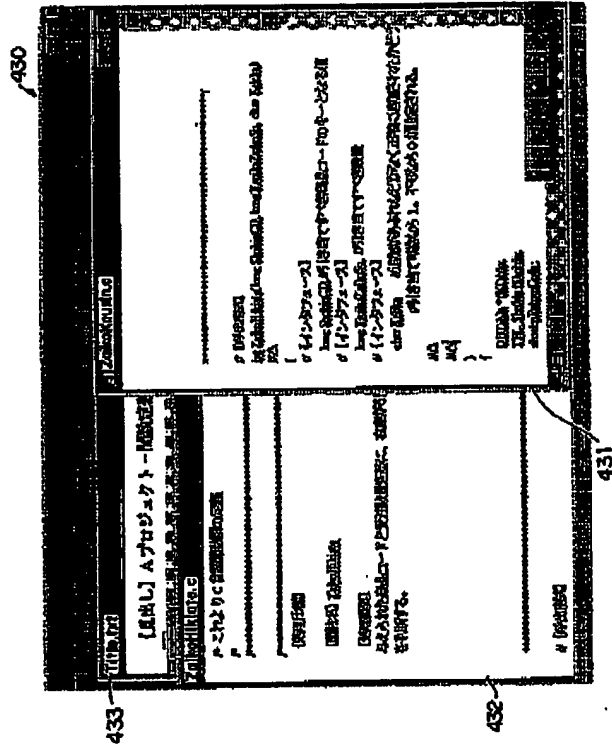
【図21】

文書編集システムを実行する
コンピュータのハードウェア構成を示す図



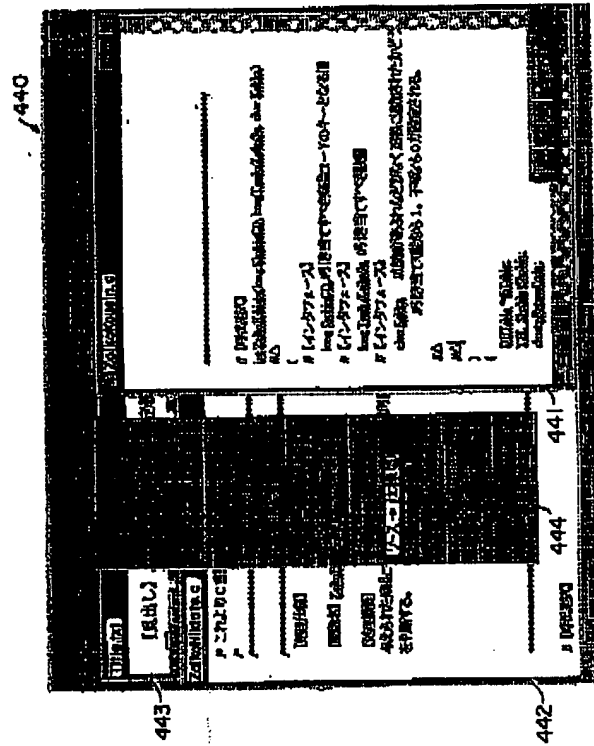
【図 15】

選択された抽出元の文書の内容を
表示する画面の例を示す図



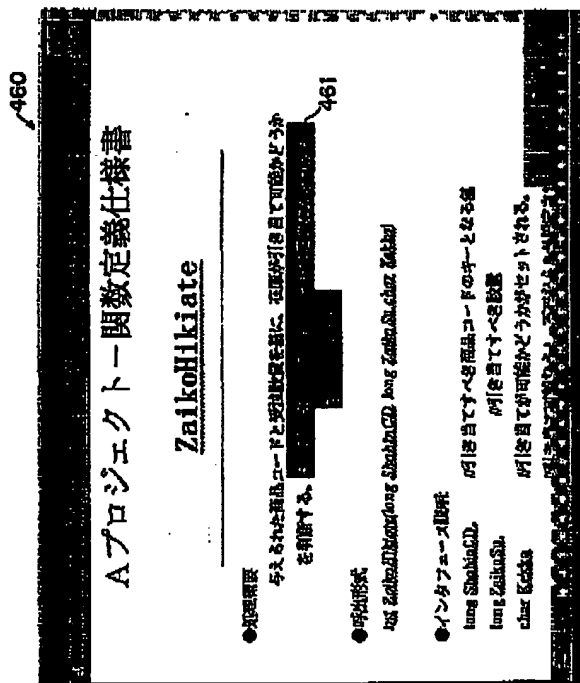
【図 16】

再構成処理を指示する画面の例を示す図



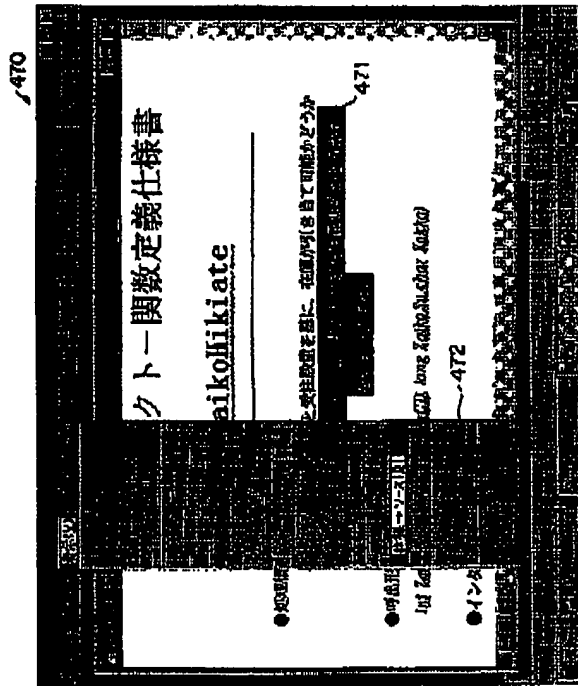
【圖 18】

表示された再構成結果に対して
編集を行った画面の例を示す図



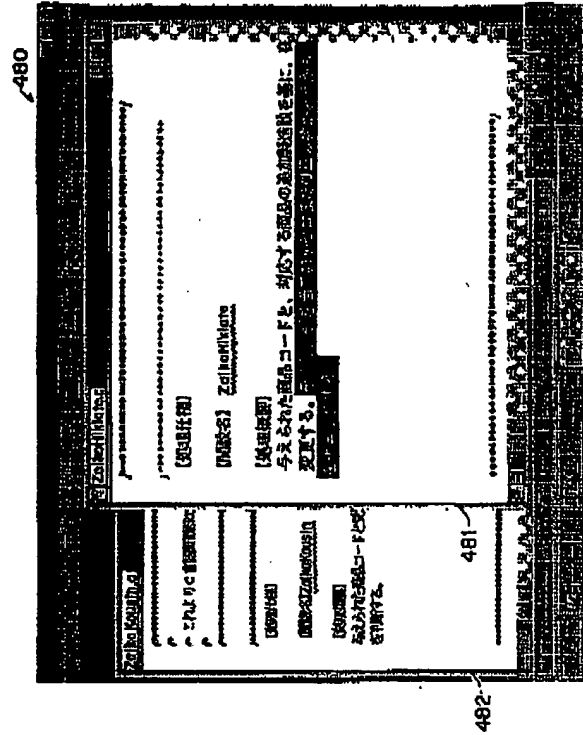
【図19】

編集結果を抽出元の文書に
 抽出する旨指示する画面の例を示す図



【図20】

編集結果が反映された抽出元の
 文書の内容を示す画面の例を示す図



**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☒ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER: _____**

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox